

Dr. Pardeep Kumar

JAYPEE UNIVERSITY OF INFORMATION TECHNOLOGY, WAKNAGHAT

TEST -2 EXAMINATION- 2019

B.Tech VIII Semester

COURSE CODE: 15B1WCI833

MAX. MARKS: 25

COURSE NAME: Big Data Analytics

COURSE CREDITS: 03

MAX. TIME: 1Hr 30 Min

Note: All questions are compulsory. Carrying of mobile phone during examinations will be treated as case of unfair means.

1. [CO2] Explain map reduce workflow with suitable diagram. How word count problem can be solved with map reduce logic? Explain with suitable diagram. [5]
2. [CO2] Name the open source implementation of map-reduce framework. Why Apache Spark is faster than Apache Hadoop? Explain HDFS file system with suitable diagram [5]
3. [CO3] Explain stream processing framework with suitable diagram. What are the problems to process the stream data? How do you differentiate the standing queries from ad hoc one? Differentiate with suitable examples. [5]
4. [CO3] What do you understand by stream data? Give examples. Compare the sampling of a fixed proportion with sampling of a fixed-size sample for sampling from a data stream. Prove reservoir sampling with principal of mathematical induction. [5]
5. [CO3] Prove that the space complexity of DGIM method which is used to count the number of 1s in the recent stream tuples is $O(\log^2 N)$. Exercise DGIM method to count the number of 1s in the stream 101010000001010100000100001111100001111100000010101010101. [5]